Acta Physica Austriaca, Suppl. XXII, 531–586 (1980)

CONFINEMENT AND TOPOLOGY IN NON-ABELIAN GAUGE THEORIES[+]

by

G. 't HOOFT
Institute for Theoretical Physics
Princetonplein 5, P.O.Box 80006
3508 TA Utrecht, The Netherlands

ABSTRACT

Pure non-Abelian theories with gauge group SU(N) are considered in 3 and 4 space-time dimensions. In 3 dimensions non-perturbative features invalidate ordinary coupling constant expansions. Disorder operators can be defined in 3 and 4 dimensions. Confinement is first explained in 3 dimensions as a spontaneous breakdown of a topologically defined Z(N) global symmetry of the theory. In 4 dimensions confinement can be seen as one of the various possible phases of the system by considering a box with periodic boundary conditions in the "thermodynamic limit". An exact duality equation allows either electric or magnetic flux tubes to be stable, but not both.

Special attention is given to the explicit oc-

[+] Lecture given at XIX. Internationale Universitätswochen für Kernphysik, Schladming, February 20 - 29, 1980.

currence of instanton configurations with an instanton angle θ in deriving duality.

I. INTRODUCTION

The first non-Abelian gauge theory that was recognized to describe interactions between elementary particles, to some extent, was the Weinberg-Salam-Ward-GIM model[1]. That model contains not only gauge fields but also a scalar field doublet H. Perturbation expansion was considered not about the point H = 0, but about the "vacuum value"

$$H = \binom{F}{0} .$$

Such a theory is usually called a theory with "spontaneous symmetry breakdown" [2]. In contrast one might consider "unbroken gauge theories" where perturbation expansion is only performed about a symmetric "vacuum". These theories are characterized by the absence of a mass term for the gauge vector bosons in the Lagrangian. The physical consequences of that are quite serious. The propagators now have their poles at $k^2=o$ and it will often happen that in the diagrams new divergences arise because such poles tend to coincide. These are fundamental infrared divergencies that imply a blow-up of the interactions at large distance scales. Often they make it nearly impossible to understand what the stable particle states are.

A particular example of such a system is "Quantum Chromodynamics", an unbroken gauge theory with gauge group SU(3), and in addition some fermions in the 3-representation of the group, called "quarks". We will investigate the possibility that these quarks are permanently confined inside bound structures that do not carry gauge quantum numbers. First of all this idea is not as absurd as it may seem. The converse would be equally difficult to understand. Gauge quantum numbers are a priori only defined up to local gauge transformations. The existence of *global* quantum numbers that would correspond to these local ones but would be detectable experimentally from a distance is not at all a prerequisite. We are nevertheless accustomed to attaching a global significance to local gauge transformation properties because we are familiar with the theories with spontaneous breakdown. The electron and its neutrino, for example, are usually said to form a gauge doublet, to be subjected to local gauge transformations. But actually these words are not properly used. Even the words "spontaneous breakdown" are formally not correct for local gauge theories (which is why I put them between quotation marks). The vacuum *never* breaks local gauge invariance because it itself is gauge invariant. All states in the physical Hilbert space are gauge-invariant. This may be confusing so let me illustrate what I mean by considering the familiar Weinberg et al model. The invariant Lagrangian is

$$L^{inv} = -\frac{1}{4}G^a_{\mu\nu}G^a_{\mu\nu}-\frac{1}{4}F_{\mu\nu}F_{\mu\nu}-D_\mu H^* D_\mu H-V(|H|)$$
$$- \bar{\psi}_L\gamma D\psi_L-\bar{e}_R\gamma De_R-k\bar{e}_R(H^*\psi_L)-k(\bar{\psi}_L H)e_R \; . \qquad (1.1)$$

Here H isthe scalar Higgs doublet. The gauge group is SU(2)xU(1), to which correspond $A^a_\mu(G^a_{\mu\nu})$ and $A^0_\mu(F_{\mu\nu})$ The subscripts L and R denote left and right handed components of a Dirac field, obtained by the projection operators $\frac{1}{2}(1 \pm \gamma_5)$.

e_R is a singlet;

ψ_L is a doublet.

D_μ stands for covariant derivative.

The function $V(|H|)$ takes its minimum at $|H|=F$. Usually one takes

$$\langle H\rangle_{\text{vacuum}} = \binom{F}{0} , \tag{1.2}$$

and perturbes around that value: $H = \binom{F + h_1}{h_2}$.

One identifies the components of ψ_L with neutrino and electron:

$$\psi_L = \binom{\nu_L}{e_L} . \tag{1.3}$$

However, this model is *not* fundamentally different from a model with "permanent confinement". One could interpret the same physical particles as being all gauge singlets, bound states of the fundamental fields with extremely strong confining forces, due to the gauge fields A_μ^a of the group SU(2). We have scalar quarks (the Higgs field H) and fermionic quarks (the ψ_L field) both as fundamental doublets. Let us call them q. Then there are "mesons" ($q\bar{q}$) and "baryons" (qq). The neutrino is a "meson". Its field is the composite, SU(2)-invariant

$$H^* \psi_L = F \ \nu_L + \text{negligible higher order terms.}$$

The e_L field is a "baryon", created by the SU(2)-invariant

$$\varepsilon_{ij} H_i \psi_{Lj} = Fe_L + \ldots \quad , \tag{1.4}$$

the e_R field remains an SU(2) singlet.
Also bound states with angular momentum occur: The neutral intermediate vector boson is the "meson"

$$H^* D_\mu H = \tfrac{i}{2} g F^2 A_\mu^{(3)} + \text{total derivative} + \text{higher orders}, \tag{1.5}$$

if we split off the total derivative term (which corresponds to a spin-zero Higgs particle).
The $W^\pm$ are obtained from the "baryon" $\varepsilon_{ij} H_i D_\mu H_j$ and the Higgs particle can also be obtained from $H^* H$
Apparently some mesonic and baryonic bound states survive perturbation expansion, most do not (only those containing a Higgs "quark" may survive).

Is there no fundamental difference then between a theory with spontaneous breakdown and a theory with confinement? Sometimes there is. In the above example the Higgs field was a faithful representation of SU(2). This is why the above procedure worked. But suppose that all scalar fields present were invariant under the center Z(N) of the gauge group SU(N), but some fermion fields were not. Then there are clearly two possibilites. The gauge symmetry is "broken" if physical objects exist that transform non-trivially under Z_N, such as the fundamental fermions. We call this the Higgs phase. If on the other hand all physical objects are invariant under Z_N, such as the mesons and the baryons, then we have permanent confinement.

Quantum Chromodynamics is such a theory where these distinct possibilities exist. It is unlikely that one will ever prove from first principles that permanent confinement takes place, simply because one can always imagine the Higgs mode to occur. If no fundamental scalar fields exist then one could introduce composite fields such as

$$H_{ab} = G^{a}_{\mu\nu} G^{b}_{\mu\nu} \quad ,$$

or

$$H^{j}_{i} = \bar{\psi}_{i} \psi^{j} \quad ,$$

and postulate nonvanishing vacuum expectation values for them:

$$\langle H_{ab} \rangle = F_1 d_{ab8} + F_2 d_{ab3}$$

or
$$\langle H^{j}_{i} \rangle = F_1 \lambda^{j}_{8i} + F_2 \lambda^{j}_{3i} \quad .$$

In that case there would be no confinement. Whether or not $F_{1,2}$ are equal to zero will depend on details of the dynamics. Therefore, dynamics must be an ingredient of the confinement mechanism, not only topological arguments. What we will attempt in this lecture is to show that topological arguments imply for this theory the existence of phase regions, separated by sharp phase transition boundaries (usually of first order). One region corresponds to what is usually called "spontaneous breakdown", and will be referred to as Higgs phase. Another corresponds to absolute quark confinement. Still another phase exists which allows for long range Coulomb-like forces to occur. (Coulomb phase.)

It is illustrative to consider first pure gauge theories in 3 space-time dimensions. These differ in two important ways from their 4 dimensional counter parts. First, they are not scale-invariant in the classical limit. A consequence of that is that they do not have a computable small coupling constant expansion. Already at finite orders in the coupling constant g phenomena occur that are associated to a compex sort of vacuum instability. This is explained in Sects 2 and 3. Secondly, the topological properties are different. In three dimensions a "disorder parameter", being an operator-valued field $\Phi(\vec{x},t)$ can be defined. If

$$\langle \Phi(x,t) \rangle_{\text{vacuum}} \neq 0 \quad ,$$

then there is absolute confinement, as we explain in Sects. 4 and 5. The remaining sections are devoted to the four-dimensional case. They overlap to some extent lectures given at Cargèse[3] except for a more explicit consideration of effects due to instantons and their angle θ. Here it is convenient to introduce the "periodic box" (a cubic or rectangular box with periodic or pseudo-periodic boundary conditions). Again we have order and disorder operators but now they are defined not on space-time points (x,t) but on loops (C,t). In Sect. 7 it is explained how to interpret these operators in terms of magnetic and electric flux operators, and magnetic flux is defined in terms of the boundary conditions of the box. There are also other conserved quantities (Sect.8), to be interpreted as electric flux ard instanton angle.

In Sect. 9 we consider the "hot box", a box at finite temperature $T = 1/k\beta$, and express the free energy F of such a box in terms of functional integrals with twisted periodic boundary conditions in 4-dimensional Euclidean space. In Sect. 10 we notice an exact duality relation fot the free energy of electric and magnetic fluxes. Section 11 shows that if electric confinement is assumed, magnetic confinement is excluded, and vice versa. The Coulomb phase realized for instance in the Georgi-Glashow model [4] is dually symmetric, as explained in sect. 12.

II. DIFFICULTIES IN THE PERTURBATION EXPANSION FOR QCD IN 2+1 DIMENSIONS

Consider pure gauge theory in 2+1 dimensions. The Lagrangian is

$$L = -\frac{1}{4} G^a_{\mu\nu} G^a_{\mu\nu} , \tag{2.1}$$

with

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu . \tag{2.2}$$

The functional integrals to be studied are

$$\int DA \exp(i \int L \, d^2\vec{x} \, dt + \text{source terms}) . \tag{2.3}$$

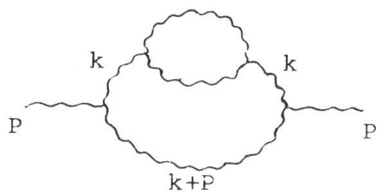
Clearly $\int L \, d^2x \, dt$ has to be dimensionless, therefore A^a_μ has dimension $(\text{mass})^{\frac{1}{2}}$ and g has dimension $(\text{mass})^{\frac{1}{2}}$.

Gauge-invariant quantities are ultra-violet convergent if they are regularized in a gauge invariant way (for instance by dimensional regularization). This is because the only possible counter terms would be

$$g^2 G_{\mu\nu} G_{\mu\nu} ; \; g^4 G_{\mu\nu} G_{\mu\nu} , \text{ etc.} \tag{2.4}$$

Which would all be of too high dimension. Conventionally this would imply that all Green's functions in Euclidean space would be well defined perturbatively. The physical theory (in Minkowsky space) would then be obtained by analytic continuation.

However, in our case we do have an infra-red problem, even in Euclidean space. Consider namely diagrams of the following type:



(2.5)

Simple power counting tells us that the small self-energy insertion is proportional to $g^2|k|$, where k is the momentum circulating in the large blob. Because of the two propagators the k integration has an infra-red divergent part:

$$\int d^3k \frac{|k|}{(k^2-i\varepsilon)^2} \text{ (remainder)}. \qquad (2.6)$$

Here the divergence is only logarithmic, but it becomes worse if more self-energy insertions occur in the k-propagator. How should one cure such divergences?

To understand the physics of this infra-red divergence let us consider on easier case first: $g\phi^3$ theory without mass term:

$$L = -\frac{1}{2}(\partial_\mu\phi)^2 - \frac{g}{3!}\phi^3 , \qquad (2.7)$$

this time in four space-time dimensions: Power counting tells us that ϕ has dimension (mass) [1] and g has dimension (mass) [1]. This theory is a little bit more artifical because an ultraviolet divergence has to be subtracted by a mass-counter term:

$$\Delta L \propto g^2 \phi^2 , \tag{2.8}$$

but still we could consider starting perturbation theory with no residual mass. If we take the same diagram (2.5) then now the self-energy insertion behaves as

$$g^2 \log |k| , \tag{2.9}$$

and the k integration is again infra-red divergent:

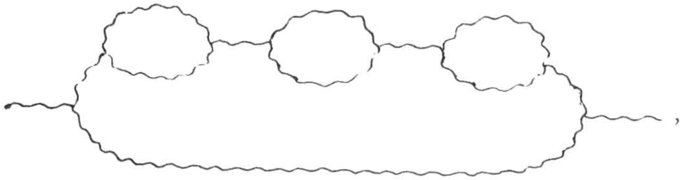
$$\int d^4k \frac{\log |k|}{(k^2 - i\varepsilon)2} \text{(remainder)} . \tag{2.10}$$

In this case however the cure seems obvious: we were doing perturbation expansion at a very singular point. No problem arises if we first introduce a small mass term

$$L \to L - \frac{1}{2} \mu^2 \phi^2 , \tag{2.11}$$

and then let $\mu^2 \downarrow 0$ in the end. How are the infinities such as (2.10) "regularized" if we do that?

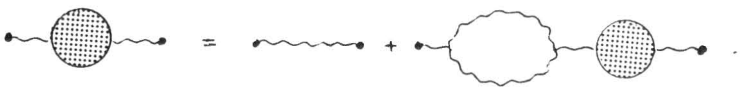
If k^2 in (2.10) is replaced by $k^2 + \mu^2$ then the limit $\mu^2 \downarrow 0$ does not exist. However one may argue that this infinity is not very physical. Suppose we sum diagrams of the type:

 (2.12)

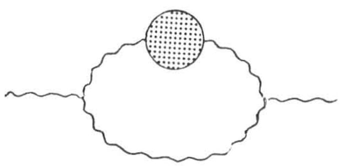
then the k integration is of the form

$$\int d^4k \frac{1}{k^2+a_1g^2\log|k|+\mu^2-i\varepsilon} \text{(remainder)}. \quad (2.13)$$

Clearly the limit $\mu^2\downarrow 0$ exist here. It has a small imaginary part due to a tachyonic pole in the dressed propagator, which cannot be admitted in a real theory, but this is probably due to the fact that ϕ^3 theory is unstable and we ignore it. The poit is that the "summed" theory is supposed to be free of infra-red divergences. In writing down (2.13) we replaced a sometimes divergent sum of bubles by an analytic and convergent expression. This expression could have been obtained directly if we wrote the Dyson-Schwinger equations for these diagrams:

 (2.14)

If eq. (2.14) is used then the diagram

 (2.15)

yields (2.13). Now (2,14) is only a truncated Dyson-Schwinger equation. If we would use the complete set of equations we would get the complete amplitude and this may be assumed to be free of divergences. (Also, in a good theory, free of tachyonic poles or cuts). A good theory should be a solution to its Dyson-Schwinger equations with dressed propagators not more singular than $1/k^2$.

Let us return to pure gauge theory in 2+1 dimensions. We can now roughly estimate how the integral (2.6) has to be cut-off, by replacing the bare propagators by dressed ones. This is not exactly according to Dyson-Schwinger equations but good enough for our purpose. The dressed propagator is

$$\frac{1}{k^2}\, f\,(g^2/|k|) \quad . \tag{2.16}$$

Let us assume that $f(z)$ has an expansion $1+a_1 z + a_2 z^2 + \ldots$ but is non singular at $z \to \infty$.
(This asymptotic expansion for f is not quite right, as as we will see later). Our integral is

$$I = \int d^3k \, \frac{f(g^2/\,|k|)}{k^2}\, R(k) \ , \tag{2.17}$$

where $R(k)$ is the non-singular remainder. Let us expand R:

$$R(k) = R_o + R_\mu\, k_\mu + R_{\mu\nu}\, k_\mu\, k_\nu + \ldots , \tag{2.18}$$

and split the integral in two pieces:

$$\int d^3k = \int_{|k|<\varepsilon} d^3k + \int_{|k|>\varepsilon} d^3k \ , \tag{2.19}$$

with $g^2 << \varepsilon << 1$. We write (2.19) as

$$I(g^2) = I_1(\varepsilon, g^2) + I_2(\varepsilon, g^2) \quad . \tag{2.20}$$

It is easy now to establish how $I_{1,2}$ behave for small ε and g^2:

$$\begin{aligned} I_2(\varepsilon, g^2) &= \\ &\int_{|k|>\varepsilon} \frac{d^3k}{k^2} \left(1 + \frac{a_1 g^2}{|k|} + \frac{a_2 g^4}{|k|^2} + \ldots\right)(R_o + R_\mu k_\mu + R_{\mu\nu} k_\mu k_\nu \ldots) \\ &= \sum_n A_n (g^2)^n - 4\pi R_o \left(\varepsilon + g^2 a_1 \log \varepsilon - \frac{g^4 a_2}{\varepsilon} \ldots\right) \\ &- \frac{4\pi}{3} R_{\mu\mu} \left(\frac{\varepsilon^3}{3} + \frac{g^2 a_1 \varepsilon^2}{2} + g^4 a_2 \varepsilon + g^6 a_3 \log \varepsilon \ldots\right) \end{aligned} \tag{2.21}$$

And if we write $k_\mu = g^2 \omega_\mu$, then

$$\begin{aligned} I_1(\varepsilon, g^2) &= \\ &g^2 \int_{|\omega|<\varepsilon/g^2} \frac{d^3\omega}{\omega^2} \left(1 + \frac{a_1}{\omega} + \frac{a_2}{\omega^2} + \ldots\right)(R_o + g^2 R_\mu \omega_\mu + g^4 R_{\mu\nu} \omega_\mu \omega_\nu \ldots) \\ &= \sum_n B_n (g^2)^n + 4\pi g^2 R_o \left(\frac{\varepsilon}{g^2} + a_1 \log \frac{\varepsilon}{g^2} + \ldots\right) \\ &+ \frac{4\pi}{3} g^6 R_{\mu\mu} \left(\frac{\varepsilon^3}{3g^6} + a_1 \frac{\varepsilon^2}{2g^4} + a_2 \frac{\varepsilon}{g^2} + a_3 \log \frac{\varepsilon}{g^2} + \ldots\right) \end{aligned} \tag{2.22}$$

Taking the two series together we find that, of course, the ε dependence disappears, but the asymptotic g^2-expansion does not only contain powers of g^2:

$$I(g^2) = \sum_n (A_n+B_n)(g^2)^n - 4\pi g^2 R_o a_1 \log g^2 - \frac{4\pi}{3} g^6 R_{\mu\mu} a_3 \log g^2 + \dots \quad (2.23)$$

So now we found where the infra-red infinity of the integral (2.6) goes: $I(g^2)$ has no proper Taylor expansion in g^2. Eq. (2.23) shows terms with log g^2. One consequence of that is that $f(g^2/_{|k|})$ in (2.16) will also develop logarithms further down in its expansion series. If we correct the previous analysis for this we find higher powers of logaritms further down the series.

Since a_1 and R_o are known, the coefficient in front of the logarithm is well determined. On the other hand, the coefficient B_1 can only be determined if f(z) is comletely known, because it must be integrated over. In general, at any fixed power of g^2, only the leading power of log g^2 has a well determined coefficient. All other coefficients can only be evaluated if the complete non-perturbative solution of f(z) is known. At higher orders also three- and more-point Green functions are required. Thus, the coefficients for the non-leading powers of log g^2 can never be determined by perturbative means alone.

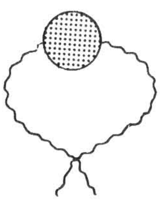
III. VACUUM STRUCTURE IN THE PERTURBATIVE 2+1 DIMENSIONAL THEORY. GAUGE INVARIANCE

Does the disease observed in the previous section have anything to do with confinement? One might argue that the dressed propagator does not exist in a theory with confinement. However, since it is not gauge-invariant, it might exist in certain convenient gauges, and we do not have to worry about spurious poles or cuts in this propagator because they need not correspond to physical states. In fact there are reasons to believe that in certain covariant gauges the propagator is actually very smooth and convergent.

Confinement is a property of the vacuum in the field theory and, as we will argue now, it is the vacuum that determines our unknown coefficients. The first unknown coefficient was B_1 in (2.22). It is

$$B_1 = \int_{|\omega|<1} \frac{d^3\omega}{\omega^2} f(1/|\omega|)\, R_o \, . \qquad (3.1)$$

Although unknown, its dependence on the remainder of the graph ,R, is very simple: it only depends on R (k=o). This implies that once it has been determined for one graph, it will be known for all other graphs as well. Let us take the diagram with R=1:



(3.2)

It corresponds to the computation of

$$\langle A^2(x)\rangle_{\text{vacuum}}$$

Of course it is ultra-violet divergent, but that divergence can be assumed to be subtracted in the usual way. We conclude that if we assume some value for

$$C_o = \langle A^2(x)\rangle^{\text{subtr.}}_{\text{vacuum}} \tag{3.3}$$

then all amplitudes can be expanded up to

$$\Gamma(k,\ldots,g^2) = \Gamma_o(k,\ldots) + g^2\Gamma_1(k,\ldots) + g^4\Gamma_2^o(k,\ldots) + g^4 \log g^2\ \Gamma_2^1(k,\ldots) + O(g^6) \quad . \tag{3.4}$$

In particular, Γ_2^o depends on C_o. At higher orders one will need

$$C_2 = \langle A^4(x)\rangle^{\text{subtr.}}_{\text{vacuum}} \;,$$

$$C_3 = \langle \partial A\ \partial A\rangle^{\text{subtr.}}_{\text{vacuum}} \;, \tag{3.5}$$

etc.

One may observe that A^2 in (3.3) is not gauge-invariant. Therefore, the actual value of C_o should not affect the Γ_2^o of gauge-invariant Green functions. Indeed, gauge-invariant Green functions have $R_o = o$, as a consequence of certain Ward-Takahashi identities.

But of course, the relevant quantities are

$$a_1 = \langle G_{\mu\nu}\ G_{\mu\nu}\rangle^{\text{subtr.}}_{\text{vacuum}}$$

$$b_1 = \langle\bar{\psi}\psi\rangle^{\text{subtr.}}_{\text{vacuum}}$$

$$a_2 = \langle(D_\alpha G_{\mu\nu})^2\rangle^{\text{subtr.}}_{\text{vacuum}} \quad \text{etc.} \tag{3.6}$$

By power counting it is possible to tell where in the power series these coefficiens enter. a_1 has dimension 3; b_1 has dimension 2; a_2 has dimension 5. And g^2 has dimension 1. Therefore, a will enter at order g^8; b_1 at order g^6 and a_2 at order g^{12}.

One may hope that iterative procedures exist to determine these coefficients semi-perturbatively: determine the known coefficients for f(z); extrapolate to find a decent function for 2 → ∞ (Padé?), then substitute this f(z) to find the next coefficients and repeat. Whatever one does, the procedure will not be as straightforward as the determination of <H> from a Higgs Lagrangian. The unknows this time form an infinite series. They are the vacuum experation values of all composite operators.

It may seem that these vacuum experation values are only needed in the 2+1 dimensional theory, and that the problems discussed here are actually irrelevant for the real world which is in 3+1 dimensions. This, we emphasize, is not true. In 3+1 dimensions the same problem occurs, however not within the ordinary perturbation expansion. This perturbation expansion namely diverges at high orders. If one tries to rearrange the perturbation expansion to obtain convergent expressions, one finds that amplitudes can be conveniently expressed by a Borel formula:

$$\Gamma(g^2) = \int_0^\infty B(z)\, e^{-z/g^2}\, dz/g^2 \quad .$$

According to the renormalization group B(z) has dimension proportional to z. If this dimension coincides with that of $\langle G_{\mu\nu} \; G_{\mu\nu} \rangle$, etc. singularities arise and their treatment goes along the same lines as in the previous section. We refer to refs. [5,6] for further details.

IV. 2+1 DIMENSIONS NON PERTURBATIVE: TOPOLOGICAL OPERATORS AND THEIR GREEN FUNCTIONS

In this section a new field is introduced that will enable us to get a better understanding of the possible vacuum structures in 2+1 dimensions. To set up our arguments step by step we begin with adding a Higgs field[7]. The gauge group is SU(N) and the gauge symmetry is spontaneously and completely broken. The Higgs fields, H, must be a set of unique representations of SU(N)/Z(N) such as the octet and decuplet representations of SU(3)/Z(3). Thus all (vector and scalar) fields are invariant under the center Z(N) of the gauge group SU(N). (This is the subgroup of matrices $e^{2\pi\, in/N}\, I$, where n is integer.) Quarks, which are not invariant under Z(N), are not yet introduced at this stage.

Besides the massive photons and Higgs particle(s) this model contains one other class of particles:

extended soliton solutions that are stable because of a topological conservation law[+]. Consider, therefore, a region R in two-dimensional space surrounded by another region B where the energy density is zero (vacuum). In B, the Higgs field $H(\vec{x})$ satisfies

$$|\langle H(\vec{x})\rangle| = F , \tag{4.1}$$

where F is a fixed number. There must be a gauge rotation $\Omega(\vec{x})$ so that

$$\Omega(\vec{x})H(\vec{x}) = H_o , \tag{4.2}$$

where H_o is fixed. $\Omega(\vec{x})$ is determined up to elements of Z(N). We also require absence of singularities so $\Omega(\vec{x})$ is continuous. Consider a closed contour $C(\theta)$ in B parametrized by an angle θ with $0 \leq \theta \leq 2\pi$ and $C(0) = C(2\pi)$. Consider the case that C goes clockwise around R. Since B is not simply connected we may have

$$\Omega(2\pi) = e^{2\pi in/N}\,\Omega(0) , \tag{4.3}$$

with $0 \leq n \leq N$, n integer. Because of continuity, n is

---

[+]For a general introduction to solitons see e.g. Coleman [8]. For an introduction to this procedure see 't Hooft [9].

conserved. If $n \neq 0$ and if we require absence of singularities in R then the field configuration in R cannot be that of the vacuum or a gauge rotation thereof, so there must be some finite amount of energy in R. The field configuration with lowest energy E, in the case $n = 1$, describes a stable soliton with mass M=E. If $N > 2$ then solitons differ from antisolitons (which correspond to n=N-1) and the number of solitons minus antisolitons is conserved modulo N.

An alternative way to represent the fields corresponding to a soliton configuration is to extend $\Omega(\vec{x})$ to be also within R (which, however, may be possible only if we admit a singularity for Ω at some point $\vec{x}_o$ in R). We then apply Ω^{-1} to the above field configuration. This has the advantage that everywhere in B we keep $H=H_o$, regardless of the number n of solitons in R, but the price of that is to allow for a singularity $\vec{x}_o$ in R. We will refer to this as the "second representation" of the soliton, for later use.

A set of operators $\Phi(\vec{x})$ is now defined as follows. Let $|A_i(\vec{x}), H(\vec{x})>$ be a state in Hilbert space which is an eigenstate of the space components of the vector fields and the Higgs fields, with $A_i(\vec{x})$ and $H(\vec{x})$ as given eigenvalues. Then

$$\Phi(\vec{x}_o)|A_i(\vec{x}), H(\vec{x})> = |A_i^{\Omega[\vec{x}_o]}(\vec{x}), H^{\Omega[\vec{x}_o]}(\vec{x})> , \qquad (4.4)$$

where $\Omega^{[\vec{x}_o]}$ gauge rotation with the property that for every closed curve $C(\theta)$ that encloses $\vec{x}_o$ once we have

$$\Omega^{[\vec{x}_o]}(\theta=2\pi) = e^{\pm 2\pi i/N}\,\Omega^{[\vec{x}_o]}(\theta=0) , \qquad (4.5)$$

where the minus sign holds for clockwise and the + sign for anticlockwise C. When $\vec{x}_o$ is outside C, then

$$\Omega^{[\vec{x}_o]}(\theta=2\pi) = {}^{[\vec{x}_o]}(\theta=0) . \tag{4.6}$$

The singularity of $\Omega^{[\vec{x}_o]}$ at $\vec{x}=\vec{x}_o$ must be smeared over an infinitesimal region around $\vec{x}_o$ but we will not consider this "renormalization" problem in this paper. In what sense is $\Phi(\vec{x})$ a *local* operator? The operator formalism in gauge theories is most conveniently formulated in the gauge $A_o(\vec{x},t)=0$. Then, time-independent continuous gauge rotations $\Omega(\vec{x})$ still form an invariance group. For physical states $|\psi\rangle$ however,

$$\langle AH|\psi\rangle = \langle A^{\Omega}\ H^{\Omega}|\psi\rangle \tag{4.7}$$

where Ω is any single-valued gauge rotation. So $\Phi(\vec{x})$ would have been trivial were it not that $\Omega^{[\vec{x}_o]}$ has a singularity at $\vec{x}_o$.

The operator $\Phi(\vec{x})$ leads physical states into physical states, and the details of Ω apart from (4.5) and (4.6) are irrelevant. It is now easy to verify that

$$\begin{aligned} \Phi(\vec{x})\Phi(\vec{y})|\psi\rangle &= \Phi(\vec{y})\Phi(\vec{x})|\psi\rangle , \\ \Phi^{+}(\vec{x})\Phi(\vec{y})|\psi\rangle &= \Phi(\vec{y})\Phi^{+}(\vec{x})|\psi\rangle, \end{aligned} \tag{4.8}$$

if $|\psi\rangle$ is physical state, because both the left- and the right-hand sides of (4.8) are completely defined by the singularities alone of the combined gauge rotations. Also, when $R(\vec{x})$ is a conventional gauge-invariant field operator composed of fields at $\vec{x}$ then obviously

$$[R(\vec{x}),\Phi(\vec{y})] = 0 \text{ for } \vec{x} \neq \vec{y}, \quad \text{but not necessarily for } \vec{x} = \vec{y} \tag{4.9}$$

Because of (4.8) and (4.9) Φ is considered to be a local field operator when it acts on physical states.

From its definition it must be clear that $\Phi(\vec{x})$ absorbs one topological unit, so we say that $\Phi(\vec{x})$ is the annihilation (creation) operator for one "bare" soliton (antisoliton) at $\vec{x}$ and $\Phi^+(x)$ is the creation (annihilation) operator for one "bare" soliton (antisoliton).

Let us now illustrate how one can compute Green functions involving $\Phi(x)$ by ordinary saddle-point techniques in a functional integral. Let us consider $<T(\Phi(0,t_1)\ \Phi^+(0,0))>= f(t_1)$ by computing the corresponding functional-integral expression:

$$f(t_1) = \frac{\int_C DADH \exp S(A,H)}{\int DADH \exp S(A,H)} , \tag{4.10}$$

where C is the set of field configurations where the fields make a sudden gauge jump at t=0 described by $\Omega^{[0]}$ (see (4.5) and (4.6)) and at $t=t_1$ they jump back by a transformation $\Omega^{[0]+}$ fields must be continuous everywhere else.

This was how $f(t_1)$ follows from the definitions but it is more elegant to transform a little further. By gauge transforming back in the region $0 < t < t_1$ we get that the fields are continuous everywhere except at $\vec{x}=0, 0 < t < t_1$. So there is a Dirac string [10] going from (0,0) to (0,t). At $0 < t < t_1$ we obtain in this way the soliton in its "second representation": a non-trivial field configuration with a singularity at the origin. In short: $\langle\Phi(x)\Phi^+(0)\rangle$ is obtained by integrating over field configurations with a Dirac string in space-time from 0 to x.

Let us compute $f(t_1)$ for Euclidean time $t_1 = i\tau$, τ real. Let the field theory be pure SU(N) without Higgs scalars. We must find the least negative action configuration with the given Dirac string. The conditions (4.5),(4.6) can be realized for an Abelian subgroup of gauge rotations Ω, so let us take

$$\Omega^{[0]^+}(\theta) = \begin{bmatrix} e^{-i\theta/N} & & & \emptyset \\ & \ddots & & \\ & & e^{-i\theta/N} & \\ \emptyset & & & e^{i\theta(1-1/N)} \end{bmatrix}, \qquad (4.11)$$

the last diagonal element being different from the others because we must have $\det\Omega = 1$ for all θ. Here θ is the angle around the time axis (remember space-time is here three dimensional). The singularity at the Dirac string must be the one obtained when $\Omega^{[0]^+}$ acts on the vacuum.

The transformations (4.11) from an Abelian subgroup, so, as as ansatz, the field configuration with this string singularity may be chosen to be an Abelian subset of field corresponding to this subgroup:

$$\frac{1}{2}\lambda^a_{ij} A^a_\mu(x) = a_\mu(x)\,\lambda^{ij}(N) \;;$$

$$\lambda^{ij}_{(N)} = \begin{bmatrix} 1 & & & & \emptyset \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ \emptyset & & & & 1-N \end{bmatrix}, \qquad (4.12)$$

we have

$$L = \frac{1}{4} G^a_{\mu\nu} G^a_{\mu\nu} = -\frac{1}{2} N(N-1) F_{\mu\nu} F_{\mu\nu} ,$$

with

$$F_{\mu\nu} = \partial_\mu a_\nu - \partial_\nu a_\mu \quad . \tag{4.13}$$

Here λ^a_{ij} are the convetional Gell-Mann matrices extended to SU(N). Within this set of fields we just have the linear Maxwell equations, and the Dirac string is the one corresponding to two oppositely charged Dirac monopoles, one at 0 and one at t_1. Their magnetic charges are $\pm 2\pi/gN$.

The total action of this configuration is

$$S = \frac{1}{2} N(N-1) \int F_{\mu\nu} F_{\mu\nu} d^3x =$$

$$= - \frac{N-1}{2N} \left(\frac{4\pi^2}{g^2}\right) \left(Z_1 + Z_2 - \frac{1}{4\pi|\tau|}\right) , \tag{4.14}$$

where Z_1 and Z_2 are the self-energies of the monopoles, which diverge but can be subtracted, leaving a space-time independent renormalization constant.

We now assume that (4.14) is indeed an absolute extremum for the total action of all field configurations with the given string singularity. We think that this assumption is plausible but present no proof. We thus obtain a first approximation to (4.10):

$$f(t_1) = A \exp\left(\frac{(N-1)\pi}{2g^2|\tau|N} + O(\log(g^2\tau))\right) ,$$

$$t_1 = i\tau , \qquad \tau \text{ real} . \tag{4.15}$$

Here A is a fixed constant obtained after subtraction of the infinite self-action at the sources. Note the plus sign in our exponent due to the attractive force between the monopole-antimonopole pair. Computation of the terms of higher order in $g^2\tau$ suffers from the obstacles mentioned in sect. 2. In any case, at large expect no convergence. Of course, when the Higgs mechanism is turned on then the soliton acquires a finite mass, say M, and then at large τ we expect

$$f(t_1) \to A\exp(-M|\tau|) ,$$

$$\tau \to \infty ,$$

$$t_1 = i\tau ,$$

$$\tau = \text{real} . \qquad (4.16)$$

just as any ordinary (dressed) propagator.

However, there are also interactions, in particular the N-soliton processes, described essentially by

$$f(x_1,\ldots,x_N) = \langle T(\Phi(x_1)\Phi(x_2)\ldots\Phi(x_N))\rangle \qquad (4.17)$$

where N is the group parameter. Again we choose $\{x_k\}$ to be Euclidean. There is some freedom in choosing the Dirac strings, for instance we can let one string leave at each point $x_1,\ldots,x_{N-1}$ and let these all assemble at x_N. Because of the modulo N conservation law, the N-1 quanta coming in at x_N are equivalent to one quantum leaving at x_N.

To find the field configuration with least negative action we could try again the ansatz (4.12) but then the result is that $x_1,\ldots,x_{N-1}$ repel each other and all attract x_N, an unsymmetric and therefore unlikely result. Obviously, any field configuration where the signs in the exponents such as (4.15) are positive, corresponding to attraction, will give much larger, therefore dominating, contributions to the amplitude. So let us try to produce such a field configuration now.

Observe that pure permutations of the N spinor components are good elements of SU(N), so by pure gauge rotations we are allowed to move the unequal diagonal element of (4.11) up or down the diagonal. So let us now again choose one Dirac string leaving at each of $x_1,\dots,x_{N-1}$ and all entering at x_N, but this time all these strings have their unequal diagonal element (see (4.11)) in a different position along the diagonal. At x_N the combined rotation is again of type (4.11) as one can easily verify, so this is a more symmetric configuration. Since the gauge transformations we performed so far are all diagonal elements of SU(N) they actually form the subgroup $[U(1)]^{N-1}$ of SU(N) which is Abelian still. Let us define

$$\lambda^{ij}(k) = \delta_{ij}-N\delta_{ik}\delta_{jk}, \quad k = 1,\dots N . \tag{4.18}$$

with

$$\sum_{k=1}^{N} \lambda_{(k)} = 0 .$$

so one of these λ matrices is actually redundant. Our present ansatz is

$$\frac{1}{2} \lambda^a_{ij} A^a_\mu(x) = \sum_{k=1}^{N} A^{(k)}_\mu(x)\lambda^{ij}_{(k)} , \tag{4.19}$$

without bothering about the invariance

$$A_\mu^{(k)}(x) \to A_\mu^{(k)} + B_\mu \ , \qquad \text{all } k \ . \tag{4.20}$$

The Lagrangian is

$$L = \frac{1}{4} G^a_{\mu\nu} G^a_{\mu\nu} = -\frac{1}{2} \sum_{k,\ell} \text{Tr}\, \lambda_{(k)} \lambda_{(\ell)} F^{(k)}_{\mu\nu} F^{(\ell)}_{\mu\nu} =$$

$$= -\frac{1}{2} N(N-1) \sum_k F^{(k)}_{\mu\nu} F^{(k)}_{\mu\nu} +$$

$$+ \frac{1}{2} N \sum_{k\neq\ell} F^{(k)}_{\mu\nu} F^{(\ell)}_{\mu\nu} \ , \tag{4.21}$$

with

$$F^{(k)}_{\mu\nu} = \partial_\mu A^{(k)}_\nu - \partial_\nu A^{(k)}_\mu \ . \tag{4.22}$$

Notice the change of sign in (4.21) when fields of different index overlap. We now choose a Dirac monopole corresponding to the kth subgroup U(1) at x_k, so $A^{(k)}(x)$ is just the field of one monopole at x_k (remember that, because we stay in the Abelian subclass of fields, linear superpositions are allowed). The diagonal terms in (4.21) then only contribute to the monopole "self-energy" (more precisely: self-action) and only contribute to the already established Z factor that must be subtracted. Only the cross terms give non-trivial effects:

$$S = \frac{1}{2} \frac{4\pi^2}{g^2 N} \left((N-1) \sum_{k=1}^{N} Z_k - \frac{1}{4\pi} \sum_{k>\ell} \frac{1}{|x_k - x_\ell|} \right) \ . \tag{4.23}$$

Again we assume that this represents the least negative action configuration without presenting the complete proof.

A good check of this equation is that if N-1 points come close and the Nth stays far away then we recover (4.14) apart from overall normalization, exactly as one should expect. So we conclude:

$$f(x_1,...,x_N)=A'\exp[\frac{\pi}{2g^2N}\sum_{k>\ell}\frac{1}{|x_k-x_\ell|}+O(\log(g^2\tau))]. \tag{4.24}$$

for Euclidean $\{x_k\}$.

The above was mainly to illustrate how to compute in lowest-order perturbation expansion the Green functions that we will discuss further in sect.5. We did not prove that the field configurations we choose to expand about are really the minimal ones (which we do expect) but there is no need to elaborate on that point further here.

V. SPONTANEOUS SYMMETRY BREAKING AND CONFINEMENT

In sect.4 we found that some SU(N) gauge theories in 2+1 dimensions possess a topological quantum number, conserved modulo N, and that Green functions corresponding to exchange of this quantum may be computed. Our field behaves as a local complex scalar field (real for SU(2) in all respects. We expect that these Green functions satisfy all the usual Wightman axioms [11] except that they are more singular than usual at small distances. In fact, the Green functions as computed in sect.4 can be exactly reproduced in a theory with free scalar particles and non-polynomial sources using superpropagator techniques [12]. We leave that as an exercise for the reader. In pure SU(N) the trouble is that only the quantum corrections give some interesting structure to these Green functions. The terms of higher orders in g^2 in (4.15) and (4.24) correspond to quantum loop corrections. They have not yet been computed, to the extent that they can be computed. However, let us assume that the resulting Green functions can be computed and are roughly generated by some effective Lagrangian with necessarily stong coupling, for instance:

$$L(\Phi,\Phi^*) = -\partial_\mu \Phi^* \partial_\mu \Phi - M^2 \Phi^*\Phi$$
$$-\frac{\lambda_1}{N!}(\Phi^N + (\Phi^*)^N) - \lambda_2 V(\Phi^*\Phi) \; . \qquad (5.1)$$

Here we have assumed that the Higgs mechanism (either by some explicit or by some dynamical Higgs field) makes all fields massive, generating also a large soliton mass M. Of course we expect many possible interaction terms but only the most important ones are added in (5.1): the λ_1 term produces the N-soliton interaction and is responsible for the non-vanishing result (4.24) for the expression (4.17), the N-point function. The λ_2 term is of course also expected and is included in (5.1) for reasons that will become clear later.

Observe now a Z(N) global symmetry that leaves the Lagrangian (5.1), and the Green functions considered in sect.4 invariant:

$$\Phi \rightarrow e^{2\pi i/N}\,\Phi,$$
$$\Phi^* \rightarrow e^{-2\pi i/N}\,\Phi^* \; . \qquad (5.2)$$

This is simply the symmetry associated with the topologically conserved soliton quantum number. Modulo N conservation laws correspond to Z(N) global symmetries.

As we saw, the mass term is essentially due to the Higgs mechanism. Roughly, $M^2 \propto \mu_H^2$ where μ_H^2 is the second derivative of the Higgs potential at the origin. We can now consider either switching off the Higgs field, or just changing the sign of μ_H^2 so that $\langle H\rangle \to 0$. What happens to M^2 (the soliton mass)? If it stays positive then the physical soliton has not gone away and the topological conservation law remains valid. The symmetry of the theory is as in the Higgs mode and we define this mode to be a "dynamical Higgs mode". However, a very good possibility is that M^2 also switches its sign, so that $\langle\Phi\rangle \to F \neq 0$. The topological global Z(N) symmetry may get spontaneously broken. This mode can be recognized directly from the Green functions of sect.4. The criterion is

$$|F|^2 = \lim_{|\tau|\to\infty} f(t_1), \quad t_1 = i\tau \ . \qquad (5.3)$$

Just for amusement we might note that (4.15) indeed seems to give $F \neq 0$ but of course the quantum corrections may not be neglected and we do not know at present how to actually compute the limit (5.3).

Spontaneous breakdown of the topological Z(N) symmetry is a new phase the system may choose, depending on the dynamics. We may compare a bunch of molecules that chooses to be in a gaseous, liquid or solid phase depending on the dynamics and on the values of certain intensive parameters.

Let us study this new phase more closely. The vacuum will now have a Z(N) degeneracy, that is, an

N-fold degeneracy. Labeling these vacua by an index from 1 to N we have

$$\langle\Phi\rangle_1 = e^{2\pi i/N} \langle\Phi\rangle_2 = \ldots \quad . \tag{5.4}$$

Since the symmetry is discrete there are no Goldstone particles; all physical particles have some finite mass. Again we are able to construct a set of topologically stable objects: the Bloch walls that separate two different vacua. These vortex-like structures are stable because the vacua that surround them are stable (Bloch walls are vortex-like because our model is in 2 + 1 dimensions). The width of the Bloch wall or vortex is roughly proportional to the inverse of the lowest mass of all physical particles, and therefore finite. The Bloch wall carries a definite amount of energy per unit of length.

We will now show the relevance of the operator

$$A(C) = \mathrm{Tr}\, P \exp \oint_C ig\, A_k(\vec{x})\, dx^k \ , \tag{5.5}$$

for these Bloch walls. Here C is an arbitrary oriented contour in 2-dim. space and P stands for the path ordering of the integral. A_k^{ij} are the space components of the gauge vector field in the matrix notation. A(C) is a non-local gauge-invariant operator that does not commute with Φ . Let us explain that. As is well known, if C' is an open contour, then the operator

$$A(C',\vec{x},\vec{x}_2) = P \exp \int_{\vec{x}_1 C'}^{\vec{x}_2} ig\, A_k(\vec{x})dx^k \ ,$$

transforms under a gauge rotation Ω as

$$A^{\Omega}(C',\vec{x}_1,\vec{x}_2) = \Omega(\vec{x}_1)A(C',\vec{x}_1,\vec{x}_2)\Omega^{-1}(\vec{x}_2) \ . \qquad (5.6)$$

Now the operator $\phi(\vec{x}_o)$ was defined by a gauge transformation Ω that is multivalued when followed over a contour that encloses $\vec{x}_o$. So when we close C' to obtain C, and C encloses $\vec{x}_o$ once,

$$A(C) = \mathrm{Tr}\, A(C,\vec{x}_1,\vec{x}_1) \ , \qquad (5.7)$$

then the value of A makes a jump by a factor $\exp(\pm 2\pi i/N)$ when the operator $\Phi(\vec{x}_o)$ acts, so

$$A(C)\Phi(\vec{x}_o) = \Phi(\vec{x}_o)A(C)\exp(2\pi in/N) \ , \qquad (5.8)$$

where n counts the number of times that C winds around $\vec{x}_o$ in a clockwise fashion minus the number of times it winds around $\vec{x}_o$ anticlockwise. Eq. (5.8) is an extension of eq. (4.9) for non-local operators $A(C)$. As we will see, (5.8) can be generalized to 3+1 dimensions. Now let us interpret (5.8) in a framework where $\Phi(\vec{x})$ is diagonalized. Then, as we see, $A(C)$ is an operator that causes a jump by a factor $\exp(2\pi in/N)$ of $\Phi(\vec{x})$ for all $\vec{x}$ inside C. So $A(C)$ causes a switch from one vacuum to another vacuum within C in the case that $Z(N)$ is spontaneously broken. In other words $A(C)$ creates a "bare" Bloch wall or vortex exactly at the curve C.

Our model does not yet include quarks. Quarks are not invariant under the center $Z(N)$, so they do not admit a direct definition of $\Phi(\vec{x})$. This difficulty is to be expected when one considers the physics of the system. The vortices that were locally stable without quarks may now become locally unstable due to virtual quark-antiquark pair creation. Most authors therefore consider quark confinement to be a basic property of the glue surrounding the quarks, in wich quarks must be inserted perturbatively. Such a procedure is justified by the experimental evidence; all hadrons can be labeled according to the number and types of quarks they contain; none of them is said to be composed of an unspecifiable or infinite number of quarks. The number of gluons on the other hand can not easily be given. We shouldn't say it is zero for most hadrons, because we need the very soft gluons to provide the binding force.

How to introduce quarks at the perturbative level is further explained in ref. [13]. The outcome is that quarks are the end point of a vortex. The conventional operator

$$\bar{\psi}(\vec{x}_1)[P \exp(\int_{\vec{x}_1}^{\vec{x}_2} ig A_k(\vec{x})dx^k)]\psi(x_2) \quad , \qquad (5.9)$$

creates not only a quark pair but also a vortex in between them. This vortex is topologically stable if $\langle\Phi\rangle = F \neq 0$. If we have a configuration with N quarks then Φ makes a full rotation over 2π when it follows a closed contour around. This is why a "baryon" consisting of N quarks is not confined to anything else. Evidently, for real baryons N must be 3.

Our conclusion is as follows. In SU(N) gauge theories where all scalar fields are in representations that are invariant under the center Z(N) of SU(N) (such as octet or decuplet representations of SU(3)), there exists a non-trivial topological global

Z(N) invariance. If the Higgs mechanism breaks SU(N) completely then the vacuum is Z(N) invariant. However, we can also have spontaneous breakdown of Z(N) symmetry. If that breakdown is complete then we can have no Higgs mechanism for SU(N), because in that mode "colored" objects are permanently and completely confined by the infinitely rising linear potentials due to the Bloch-wall-vortices. We can also envisage the intermediate modes where a Higgs mechanism breaks SU(N) partly, and Z(N) is partly broken. Finally, if neither Higgs' effect, nor spontaneous breakdown of Z(N) take place, then there must be massless particles causing complicated long range interactions as we will show more explicitly for the 3+1 dimensional case. That may either correspond to a point where a higher order phase transition occurs, or to a new phase, e.g. the Coulomb or Georgi-Glashow phase, where an effective Abelian photon field survives at long distances, see sect. 12.

Eqs. (4.8) and 5.8) are the basic commutation relations satisfied by our topological fields Φ They suggest a dual relationship between A and Φ Indeed, one could start with a scalar theory exhibiting global Z(N) invariance and then define the topological operator A(C) through eq. (5.8), but it is impossible to see this way that A can be written as the ordered exponent of an integral of a vector potential, and also the gauge group SU(N) cannot be recovered. As we will explain later, the center Z(N) is more basic to this all then the complete group SU(N).

A good name for the field Φ is the "disorder parameter" [14] since it does not commute with the other, usual, fields wich have been called order parameters in solid-state physics. The fact that in the quark confinement phase the degenerate vacuum states are eigenstates of this disorder parameter shows a close analogy with the superconductor where the vacuum state is an eigenstate of the order parameter.

VI. SU(N) GAUGE THEORIES IN 3 + 1 DIMENSIONS

In the previous sections the construction of a scalar field and the successive formulation of the spontaneous breakdown of the topological Z(N) symmetry were only possible because the model was in 2 space, 1 time dimensions. Also the boundary between different but equivalent vacua can only serve as an vortex in 2+1 dimensions. It would have the topology of a sheet in 3+1 dimensions and therefore not be useful as a vortex of conserved electric flux. So in 3+1 dimensions the formulation of quark confinement must be considerably different from the 2+1 dimensional case. Nevertheless extension of our ideas to 3+1 dimensions is possible.

We concentrate on longe-range topological phenomena. One topological feature is the instanton, corresponding to a gauge field configuration with non-trivial Pontryagin or Second Chern Class number. This however has no direct implication for confinement. What is needed for confinement is something with the space-time structure of a string, i.e. a two dimensional manifold in 4 dim. space-time. Instantons are rather event-like, i.e. zero dimensional and can for instance give rise to new types of interactions that violate otherwise apparent symmetries.

As we will see, they do play a role, though be it a subtle one. A topological structure which is extended in two dimensional sheets exists in gauge theories, as has been first observed by Nielsen, Olesen [15] and Zumino [16]. They are crucial. We will exhibit them by compactifying space-time. For the instanton it had been convenient to compactify space-time to a sphere S_4 . For our purposes a hypertorus

$$S_1 \times S_1 \times S_1 \times S_1$$

is more suitable [17]. One can also consider this to be a four dimensional cubic box with periodic boundary conditions. Inside, space-time is flat. The box may be arbitrarily large. To be explicit we put a pure SU(N) gauge theory in the box (no quarks yet). Now in the continuum theory the gauge fields themselves are representations of SU(N)/Z(N), where Z(N) is the center of the group SU(N):

$$Z(N) = \{e^{2\pi in/N} I;\ n = 0,\ldots,N-1\} . \tag{6.1}$$

This is because any gauge transformation of the type (6.1) leaves $A_\mu(x)$ invariant. A consequence of this is the existence of another class of topological quantum numbers in this box besides the familiar Pontryagin number. Consider the most general possible periodic boundary condition for $A_\mu(x)$ in the box. Take first a plane $\{x_1,x_2\}$ in the 12 direction with fixed values of x_3 and x_4. One may have

$$\begin{aligned} A_\mu(a_1,x_2) &= \Omega_1(x_2)A_\mu(0,x_2) \ , \\ A_\mu(x_1,a_2) &= \Omega_2(x_1)A_\mu(x_1,0) \ . \end{aligned} \tag{6.2}$$

Here, a_1, a_2 are the periods.
ΩA_μ stands short for

$$\Omega A_\mu \, \Omega^{-1} + \frac{1}{gi} \partial_\mu \, \Omega^{-1} \quad . \tag{6.3}$$

The periodicity conditions for $\Omega_{1,2}(x)$ follow by considering (6.2) at the corners of the box:

$$\Omega_1(a_2)\Omega_2(0) = \Omega_2(a_1)\Omega_1(0)Z \ , \tag{6.4}$$

where Z is some element of Z(N).
One may now perform continuous gauge transformations on $A_\mu(x)$,

$$A_\mu(x_1,x_2) \rightarrow \Omega(x_1,x_2)A_\mu(x_1,x_2) \ , \tag{6.5}$$

where $\Omega(x_1,x_2)$ (non-periodic) can be arranged either such that $\Omega_2(x_1) \rightarrow I$ or such that $\Omega_2(x_1) \rightarrow I$, but not both, because Z in (6.4) remains invariant under (6.5) as one can easily verify. We call this element Z(1,2) because the 12 plane was chosen. By continuity Z(1,2) cannot depend on x_3 or x_4. For each $(\mu\nu)$ direction such a Z element exist, to be labeled by integers

$$n_{\mu\nu} = -n_{\nu\mu} \quad , \tag{6.6}$$

defined modulo N. Clearly this gives

$$N^{\frac{d(d-1)}{2}} = N^6 \tag{6.7}$$

topological classes of gauge field configurations. Note that these classes disappear if a field in the fundamental representation of SU(N) is added to the system (these fields would make unacceptable jumps at the boundary). Indeed, to understand quark confinement it is necessary to understand pure gauge systems without quarks first.

As we shall see, the new topological classes will imply the existence of new vacuum parameters besides the well-known instanton[18] angle θ The latter still exists in our box, and will be associated with a topological quantum number ν , an arbitrary integer.

VII. ORDER AND DISORDER LOOP INTEGRALS

To elucidate the physical significance of the topological numbers $n_{\mu\nu}$ we first concentrate on gauge field theory in a three dimensional periodic box with time running from $-\infty$ to ∞. To be specific we will choose the temporal gauge,

$$A_4 = 0 \ . \tag{7.1}$$

(this is the gauge in wich rotation towards Euclidean space is particularly elegant). Space has the topology $(S_1)^3$. There is an infinite set of homotopy classes of closed oriented curves C in this space: C may wind any number of times in each of the three principal directions. For each curve C at each time t there is a quantum mechanical operator A(C,t) defined by

$$A(C,t) = \mathrm{Tr}\ P \exp \oint_C ig\, \vec{A}(\vec{x},t)\cdot d\vec{x} \ , \tag{7.2}$$

called Wilson loop or order parameter. Here P stands for path ordering of the factors $\vec{A}(\vec{x},t)$ when the exponents are expanded. The ordering is done with respect to the matrix indices. The $\vec{A}(\vec{x},t)$ are also operators in Hilbert space, but for different $\vec{x}$, same t, all $A(\vec{x},t)$ commute with each other. By analogy with ordinary electromagnetism we say that A(C) *measures* magnetic flux *through* C, and in the same time *creates* an electric flux line *along* C. Since A(C) is gauge-invariant under purely periodic gauge transformations, our versions of magnetic and electric flux are gauge-invariant. Therefore they are not directly linked to the gauge covariant curl $G^a_{\mu\nu}(\vec{x})$.

There exists a dual analogon of A(C) wich will be called B(C) or disorder loop operator [13].C is again a closed oriented curve in $(S_1)^3$. A simple definition of B(C) could be made by postulating its equal-time commutation rules with A(C):

$$\begin{aligned} [A(C), A(C')] &= 0; \\ [B(C), B(C')] &= 0; \\ A(C)B(C') &= B(C')A(C) \exp 2\pi in/N , \end{aligned} \tag{7.3}$$

where n is the number of times C' winds around C in a certain direction. Note that n is only well defined if either C or C' is in the trivial homotopy class (that is, can be shrunk to a point by continuous deformations). Therefore, if C' is in a nontrivial class we must choose C to be in a trivial class. Since these commutation rules (7.3) determine B(C) only up to factors that commute with A and B, we could make further requirements, for instance that B(C) be a unitary operator.

An explicit definition of B(C) can be given as follows. As in sect. 4, we go to the temporal gauge, $A_0 = 0$. We then must distinguish a "large Hilbert space" H of all field configurations A(x) from a "physical Hilbert space $\tilde{H} \subset H$. This $\tilde{H}$ is defined to be the subspace of H of all gauge invariant states:

$$\tilde{H} = \{|\psi>;\ <\vec{A}(\vec{x})|\psi> = <\Omega\vec{A}(\vec{x})|\psi>\} \quad , \qquad (7.4)$$

where Ω is any infinitesimal gauge transformation in 3 dim. space. Often we will also write Ω for the corresponding rotation in **H**:

$$\tilde{H} = \{|\psi>;\ \Omega|\psi> = |\psi>,\ \Omega \text{ infinitesimal}\} \ . \qquad (7.5)$$

Now consider a pseudo-gauge transformation $\Omega^{[C']}$ defined to be a genuine gauge transformation at all points $\vec{x} \notin C'$, but singular on C'. For any closed path $x(\theta)$ with $0 < \theta < 2\pi$ twisting n times around C' we require

$$\Omega^{[C']}(x(2\pi)) = \Omega^{[C']}(x(0))e^{2\pi in/N} \ . \qquad (7.6)$$

This discontinuity is not felt by the fields $A(\vec{x},t)$ wich are invariant under Z(N). They do feel the singularity at C' however. We define B(C') as

$$\Omega^{[C']}$$

but with the singularity at C' smoothened; this corresponds to some form of regularization, and implies that the operator differs from an ordinary gauge transformation. Therefore, even for $|\psi> \varepsilon$ **H** we have

$$B(C')|\psi> \neq |\psi> \ . \qquad (7.7)$$

For any regular gauge transformation Ω we have an Ω'

such that

$$\Omega\Omega^{[C']} = \Omega^{[C']}\,\Omega' \ . \qquad (7.8)$$

Therefore, if $|\psi\rangle \,\varepsilon\, \tilde{H}$ then $B(C')|\psi\rangle \,\varepsilon\, \tilde{H}$, and $B(C')$ is gauge-invariant. We say that $B(C')$ *maesures* electric flux *through* C' and *creates* a magnetic flux line *along* C'.

We now want to find a conserved variety of Non-Abelian gauge-invariant magnetic flux in the 3-direction in the 3 dimensional periodic box. One might be temped to look for some curve C enclosing the box in the 12 direction so that A(C) maesures the flux through the box. That turns out not to work because such a flux is not guaranteed to be conserved. It is better to consider a curve C' in the 3-direction winding over the torus exactly once:

$$C' = \{\vec{x}(s), 0 \leq s \leq 1;\ \vec{x}(1) = \vec{x}(0)+(0,0,a_3)\}. \qquad (7.9)$$

$B(C')$ creates one magnetic flux line. But $B(C')$ also changes the number n_{12} into $n_{12}+1$. This is because

$$\Omega^{[C']}$$

makes a Z(N) jump according to (7.6). If $\Omega_{1,2}(\vec{x})$ in (6.2) are still defined to be continuous then Z in (6.4) changes by one unit. Clearly, n_{12} measures the number of times an operator of the type $B(C')$ has acted, i.e. the number of magnetic flux lines created. n_{12} is also conserved by continuity. We simply define

$$n_{ij} = \varepsilon_{ijk}\, m_k \ , \qquad (7.10)$$

with m_k the total magnetic flux in the k-direction. Note that $\vec{m}$ corresponds to the usual magnetic flux (apart from a numerical constant) in the Abelian case. Here, $\vec{m}$ is only defined as an integer modulo N.

VIII. NON-ABELIAN GAUGE-INVARIANT ELECTRIC FLUX IN THE BOX

As in the magnetic case, there exists no simple curve C such that the total electric flux through C, measured by B(C), corresponds to a conserved total flux through the box. We consider a curve C winding once over the torus in the 3-direction and consider the electric flux creation operator A(C). But first we must study some new conserved quantum number s.

Let $|\psi>$ be a state in the before mentioned little Hilbert space $\tilde{H}$. Then, according to eq. (7.5), $|\psi>$ is invariant under *infinitesimal* gauge transformations Ω. But we also have some non-trivial homotopy classes of gauge transformations Ω. These are the pseudoperiodic ones:

$$\begin{aligned} \Omega(a_1,x_2,x_3) &= \Omega(0,x_2,x_3)Z_1 \ , \\ \Omega(x_1,a_2,x_3) &= \Omega(x_1,0,x_3)Z_2 \ , \\ \Omega(x_1,x_2,a_3) &= \Omega(x_1,x_2,0)Z_3 \ , \\ Z_{1,2,3} &\in \text{center } Z(N) \text{ of } SU(N) \ , \end{aligned} \tag{8.1}$$

and also those Ω which are periodic but do carry a non-trivial Pontryagin number ν. A little problem arises when we try to combine these two topological features. The $Z_{1,2,3}$ can be labeled by three integers $k_{1,2,3}$ between 0 and N:

$$Z_t = e^{2\pi i k_t/N} \ . \tag{8.2}$$

But how is ν defined? The best definition is obtained if we consider a field configuration in a *four* dimensional space, obtained by multiplying the box $(S_1)^3$ with a line segment:

$$0 \leq t \leq 1 \ .$$

Now choose a boundary condition: $A(t=1)=\Omega A(t=0)$. Then, if the fields in between are continuous, then

$$P \equiv g^2 \int G_{\mu\nu} \tilde{G}_{\mu\nu} \, d^4x/32\pi^2 \qquad (8.3)$$

is uniquely determined by Ω . On S_4 this would be the integer ν. Now however, it needs not to be integer anymore because of the twists in the periodic boundary conditions for $(S_1)^3$. We find

$$P = \frac{(\vec{m}\cdot\vec{k})}{N} + \nu \quad , \qquad (8.4)$$

where ν is integer and $\vec{m}$ is the magnetization defined in the previous section. Notice that ν is only well defined if $\vec{m}$ and $\vec{k}$ are given as genuine integers, not modulo N. Taking this warning to heart, we write $\Omega[\vec{k},\nu]$ for any Ω in the homotopy class $[\vec{k},\nu]$.

Notice that not only do the $A_\mu(x)$ transform smoothly under $\Omega[\vec{k},\nu]$, since they are invariant under the Z(N) transformations of eq.(8.1), but also their boundary conditions do not change. These Ω commute therefore with the magnetic flux $\vec{m}$. If two Ω satisfy the same equation (8.1) and have the same ν, they may act differently on states of the big Hilbert space H, but since they differ only by regular gauge transformations they act identically on states in $\tilde{H}$, defined in (7.5). We may simultaneously diagonalize the Hamiltonian H, the magnetic flux $\vec{m}$, and $\Omega[k,\nu]$:

$$\Omega[\vec{k},\nu]|\psi\rangle = e^{i\omega(\vec{k},\nu)}|\psi\rangle , \tag{8.5}$$

where $\omega(\vec{k},\nu)$ are strictly conserved numbers. Now the Ω operators form a group. Defining for each Ω the number P as in (8.4) we have

$$\Omega[\vec{k}_1,p_1]\Omega[\vec{k}_2,p_2] = \Omega[\vec{k}_1+\vec{k}_2,p_1+p_2] , \tag{8.6}$$

so

$$\omega(\vec{k}_1,\nu_1) + \omega(k_2,\nu_2) = \omega(\vec{k}_1+\vec{k}_2,\nu_1+\nu_2) , \tag{8.7}$$

and

$$\omega(\vec{k} + N\vec{\ell},\nu) = \omega(\vec{k},\nu + (\vec{\ell}\cdot\vec{m})) , \tag{8.8}$$

if $\vec{\ell}$ is an integer. We find that ω must be linear in $\vec{k}$ and ν:

$$\omega(\vec{k},\nu) = \frac{2\pi}{N}(\vec{e}\cdot\vec{k}) + \frac{\theta}{N}(\vec{m}\cdot\vec{k}) + \theta\nu , \tag{8.9}$$

where e_i are integer numbers defined modulo N, and θ is the familiar instanton angle, defined to lie between O and 2π.

Now let us turn back to A(C) defined in eq.(7.2). If C is the curve considered in the beginning of this section, A(C) is not invariant under $\Omega[\vec{k},\nu]$ because

$$A(C) \to \mathrm{Tr}\,\Omega(\vec{x}_1)[P \exp \int_C ig\,\vec{A}\,d\vec{x}]\,\Omega^{-1}(\vec{x}_1 + \vec{a}_3)$$
$$= e^{-2\pi i k_3/N}\,A(C) . \tag{8.10}$$

Therefore,

$$A(C)\Omega[\vec{k},\nu]|\psi> = \Omega[\vec{k},\nu]e^{-2\pi ik_3/N}A(C)|\psi> \quad (8.11)$$

If $\Omega[\vec{k},\nu]|\psi> = e^{i\omega(\vec{k},\nu)}|\psi>$, (8.12)

and $A(C)|\psi> = |\psi'>$, (8.13)

then $\Omega[\vec{k},\nu]\ |\psi'> = e^{i\omega(\vec{k},\nu)+2\pi ik_3/N}|\psi'>$. (8.14)

Therefore A(C) increases e_3 by one unit:

$$e_3A(C)|\psi> = A(C)(e_3+1)|\psi> \quad . \quad (8.15)$$

e_3 is a good indicator for electric flux in the 3-direction, up to a constant. It is strictly conserved. However if we let θ run from 0 to 2π then $\vec{e}$ turns into $\vec{e}+\vec{m}$. It is therefore physically perhaps more approptiate to identify

$$\vec{e} + \frac{\theta}{2}\vec{m} \quad (8.16)$$

as being the total electric flux in the three directions of the box.

IX. FREE ENERGY OF A GIVEN FLUX CONFIGURATION

Again we follow ref.[3] but for completeness we add the Pontryagin winding number ν.

Let us write down the free energy F of a given state $(\vec{e},\vec{m},\theta)$ at temperature $T = 1/k\beta$:

$$e^{-\beta F} = \underset{\tilde{H}}{\mathrm{Tr}}\; P_e(\vec{e})\; P_m(\vec{m}) P_\theta(\theta) e^{-\beta H} \quad . \tag{9.1}$$

Here H is the Hamiltonian and $\tilde{H}$ the little Hilbert space. P are projection operators. $P_m(\vec{m})$ is simply defined to select a given set of $n_{ij} = \varepsilon_{ijk}\, m_k$, the three space-like indices of eq. (6.6). $P_e(\vec{e}) P_\theta(\theta)$ is defined by selecting states $|\psi>$ with

$$\Omega[\vec{k},\nu]|\psi> = e^{\frac{2\pi i}{N}(\vec{k}\cdot\vec{e}) + \frac{\theta i}{N}(\vec{m}\cdot\vec{k}) + i\theta\nu} |\psi> \quad . \tag{9.2}$$

Therefore $P_e(\vec{e})P_\theta(\theta) =$

$$\frac{1}{N^3} \sum_{k,\nu} e^{-\frac{2\pi i}{N}(\vec{k}\cdot\vec{e}) - \frac{\theta i}{N}(\vec{m}\cdot\vec{k}) - i\theta\nu}\, \Omega[\vec{k},\nu] \quad . \tag{9.3}$$

Now $e^{-\beta H}$ is the evolution operator in imaginary time direction at interval β, expressed by a functional integral over a Euclidean box with sides (a_1, a_2, a_3, β):

$$<\vec{A}_{(1)}(\vec{x})|e^{-\beta H}|\vec{A}_{(2)}(\vec{x})> = \int DAe^{S(A)} \Bigg|_{\vec{A}(\vec{x},0)=\vec{A}_{(2)}(\vec{x})}^{\vec{A}(\vec{x},\beta)=\vec{A}_{(1)}(\vec{x})} \quad . \tag{9.4}$$

We may fix the gauge for $\vec{A}_{(2)}(\vec{x})$ for instance by choosing

$$\begin{aligned} A_{(2)3}(\vec{x}) &= 0 \;, \\ A_{(2)2}(x,y,0) &= 0 \;, \\ A_{(2)1}(x,0,0) &= 0 \;. \end{aligned} \tag{9.5}$$

We already had $A_4(\vec{x},t) = 0$. Since only states in $\tilde{H}$ are considered, we insert also a projection operator $\int_{\Omega\,\varepsilon\, I} D\Omega$ were I is the trivial homotopy class.

"Trace" means that we integrate over all $A_{(1)}=A_{(2)}$ therefore we get periodic boundary conditions in the 4-direction. Insertions of $\int_{\Omega\,\varepsilon\, I} D\Omega$ means that we have periodicity up to gauge transformations, in the completely unique gauge

$$A_4(\vec{x},\beta)=A_3(\vec{x},0)=A_2(x,y,0,0)=A_1(x,0,0,0)=0. \qquad (9.6)$$

Eq. (9.3) tells us that we have to consider twisted boundary conditions in the 41, 42, 43 directions and Fourier transform:

$$e^{-\beta F(\vec{e},\vec{m},\theta,\vec{a},\beta)} = \frac{1}{N^3}\sum_{\vec{k},\nu} e^{-\frac{2\pi i}{N}(\vec{k}\cdot\vec{e})-i\theta(\nu+\frac{\vec{m}\cdot\vec{k}}{N})} W\{\vec{k},\vec{m},\nu,a_\mu\}. \qquad (9.7)$$

Here $\mathbf{W}\{\vec{k},\vec{m},\nu,a_\mu\}$ is the Euclidean functional integral with boundary conditions fixed by choosing $n_{ij} = \varepsilon_{ijk}\, m_k$; $n_{i4} = k_i$; $a_4 = \beta$, and a Pontryagin number ν. Because of the gauge choice (9.6) this functional integral must include integration over the Ω belonging to the given homotopy classes as they determine the boundary conditions such as (6.2).
The definition of W is completely Euclidean symmetric. In the next chapter I show how to make use of this symmetry with respect to rotation over 90° in Euclidean space.

X. DUALITY

The Euclidean symmetry in eq. (9.7) suggests to consider the following SO(4) rotation:

$$\begin{bmatrix} 0 & -1 & & \\ 1 & 0 & & \\ & & 0 & 1 \\ & & -1 & 0 \end{bmatrix} \quad . \tag{10.1}$$

Let us introduce a notation for the first two components of a vector:

$$\begin{aligned} x_\mu &= (\vec{x}, x_4) \ , \\ \tilde{x} &= (x_1, x_2) \ , \\ \hat{x} &= (x_2, x_1) \ . \end{aligned} \tag{10.2}$$

We have, from eq. (9.7):

$$\exp\left[-\beta F(\tilde{e}, e_3, \tilde{m}, m_3, \theta, \tilde{a}, a_3, \beta)\right] =$$

$$= \frac{1}{N^2} \sum_{\tilde{k}, \tilde{\ell}} \exp\left[\frac{2\pi i}{N}(-(\tilde{k}\cdot\tilde{e}) + (\tilde{\ell}\cdot\tilde{m})) - a_3 F(\tilde{\ell}, e_3, \tilde{k}, m_3, \theta, \hat{a}, \beta, a_3)\right] . \tag{10.3}$$

Notice that in this formula the transverse electric and magnetic fluxes are Fourier transformed and interchange positions. Notice also that, apart from a sign difference, there is a complete electric-magnetic symmetry in this expression, in spite of the fact that the definition of F in terms of W was not so symmetric. Eq. (10.3) is an exact property of our system. No approximation was made. We refer to it as "duality".

XI. LONG-DISTANCE BEHAVIOR COMPATIBLE WITH DUALITY

Eq. (10.3) shows that the instanton angle θ plays no role in duality. It does however affect the physical interpretation of $\vec{e}$ as electric flux, see (8.10). From now on we put $\theta = 0$ for simplicity, and omit it.

Let us now assume that the theory has a mass gap. No massless particles occur. Then asymptotic behavior at large distances will be approached exponentially. Then it is excluded that

$$F(\vec{e},\vec{m},\vec{a},\beta) \to 0, \quad \text{exponentially as } \vec{a},\beta \to \infty \quad ,$$

for all $\vec{e}$ and $\vec{m}$, wich would clearly contradict (10.3). This means that at least some of the flux configurations must get a large energy content as $\vec{a}$, $\beta \to \infty$. These flux lines apparently cannot spread out and because they were created along curves C it is practically inescapable that they get a total energy wich will be proportional to their length:

$$E = \lim_{\beta\to\infty} F = \rho a \quad . \tag{11.1}$$

However, duality will never enable us to determine whether it is the electric or the magnetic flux lines that behave this way. From the requirement that W in (9.7) is always positive one can deduce the impossibility of a third option, namely that only exotic combinations of electric and magnetic fluxes behave as strings (provided $\theta = 0$).

For further information we must make the physically quite plausible assumption of "factorizability":

$$F(\vec{e},\vec{m}) \to F_e(\vec{e}) + F_m(\vec{m}) \text{ if } \vec{a},\beta \to \infty \ . \tag{11.2}$$

Suppose that we have confinement in the electric domain:

$$F_e(0,0,1) \to \rho a_3 \tag{11.3}$$

where ρ is the fundamental string constant. Then we can derive from duality the behavior of $F_m(\vec{m})$.

First we improve (11.3) by applying statistical mechanics to obtain F_e for large but finite β. One obtains:

$$e^{-\beta F_e(e_1,e_2,0,\vec{a},\beta) + C(a,\beta)} =$$

$$= \sum_{n_1^\pm, n_2^\pm} \frac{1}{n_1^+! n_2^+! n_1^-! n_2^-!} \gamma_1^{n_1^+ + n_1^-} \gamma_2^{n_2^+ + n_2^-} \delta_N(n_1^+ - n_1^- - e_1)\,\delta_N(n_2^+ - n_2^- - e_2). \tag{11.4}$$

Here

$$\gamma_1 = \lambda a_2 a_3 \, e^{-\beta\rho a_1},$$

$$\gamma_2 = \lambda a_1 a_3 \, e^{-\beta\rho a_2},$$

$$\delta_N(x) = \frac{1}{N} \sum_{k=0}^{N-1} e^{2\pi i k x/N}$$

$$= \begin{cases} 1 & \text{if } x = 0 \pmod N \\ 0 & \text{if } x = \text{other integer number.} \end{cases} \tag{11.5}$$

The sum is over all nonnegative integer values of $n_i^\pm$ (the orientations $\pm$ are needed if $N \geq 3$). The γ's are Boltzmann factors associated with each string-like flux tube.

We now insert this, with (11.2), into (10.3) putting $e_3 = m_3 = 0$. One obtains

$$e^{-\beta F_m(m_1,m_2,0,\vec{a},\beta)} = C' e^{2\sum_a \gamma'_a \cos(2 m_a \pi/N)}, \tag{11.6}$$

where C' is again a constant and

$$\gamma_1' = \lambda a_1 \beta \quad e^{-\rho a_2 a_3} \quad ,$$
$$\gamma_2' = \lambda a_2 \beta \quad e^{-\rho a_1 a_3} \quad . \tag{11.7}$$

At $\beta \to \infty$ we get

$$F_m(\tilde{m}, 0, \vec{a}, \beta) \to E_m(\tilde{m}, 0, \vec{a}) = \sum_i E_i(m_i, \vec{a}) \ ,$$

with

$$E_1(m_1, \vec{a}) = 2\lambda(1 - \cos \frac{2\pi m_1}{N}) a_1 e^{-\rho a_2 a_3} \ , \tag{11.8}$$

and similarly for E_2 and E_3.

One reads off from eq. (11.8) that there will be no magnetic confinement, because if we let the box become wider the exponential factor

$$e^{-\rho a_2 a_3}$$

causes a rapid decrease of the energy of the magnetic flux. Notice the occurrence of the string constant ρ in there.

Of course we could equally well have started from the presumption that there were magnetic confinement. One then would conclude that there would be no electric confinement, because then the electric flux would have an energy given by (11.8).

XII. THE COULOMB PHASE

To see what might happen in the absence of a mass gap one could study the (first) Georgi-Glashow model [4] Here SU(2) is "broken spontaneously" into

U(1) by an isospin one Higgs field. Ordinary perturbation expansion tells us what happens in the infrared limit. There are electrically charged particles: $W^{\pm}$ (the charged vector particles). They carry *two* fundamental electric flux units ("quarks" with isospin ½ would have the fundamental flux unit $q_o = \pm \frac{1}{2} e$). There are also magnetically charged particles (monopoles, [19].They also carry two fundamental magnetic flux units:

$$g = \frac{2\pi}{q_o} = \frac{4\pi}{e} \ . \tag{12.1}$$

A given electric flux configuration of k flux units would have an energy

$$E = \frac{q_o^2 k^2 a_1}{2a_2 a_3} \ . \tag{12.2}$$

At finite β however pair creation of $W^{\pm}$ takes place, so that we should take a statistical average over various values of the flux. Flux is only rigorously defined modulo $2q_o$. We have

$$e^{-\beta F_e(1,0,0)} = \frac{\sum_{k=-\infty}^{\infty} \exp[-\beta \frac{e^2 a_1}{2a_2 a_3}(k+\frac{1}{2})^2]}{\sum_{k=-\infty}^{\infty} \exp[-\beta \frac{e^2 a_1}{2a_2 a_3} k^2]} \ . \tag{12.3}$$

Similarly, because of pair creation of magnetic monopoles

$$e^{-\beta F_m(1,0,0)} = \frac{\sum_{k=-\infty}^{\infty} \exp[-\beta \frac{8\pi^2 a_1}{e^2 a_2 a_3}(k+\frac{1}{2})^2]}{\sum_{k=-\infty}^{\infty} \exp[-\beta \frac{8\pi^2 a_1}{e^2 a_2 a_3} k^2]} \ . \tag{12.4}$$

These expressions do satisfy duality, eq. (10.3). This is easily verified when one observes that

$$\sum_{k=-\infty}^{\infty} e^{-\lambda k^2} = \sqrt{\frac{\pi}{\lambda}} \sum_{k=-\infty}^{\infty} e^{-\pi^2 k^2/\lambda} ,$$

and

$$\sum_{k=-\infty}^{\infty} (-1)^k e^{-\lambda k^2} = \sqrt{\frac{\pi}{\lambda}} \sum_{k=-\infty}^{\infty} e^{-\pi^2 (k+\frac{1}{2})^2/\lambda} . \qquad (12.5)$$

Notice now that this model realizes the dual formula in a symmetric way, contrary to the case that there is a mass gap. This dually symmetric mode will be referred to as the "Coulomb phase" or "Georgi-Glashow phase".

Suppose that Quantum Chromodynamics would be enriched with two free paramaters that would not destroy the basic topological features (for instance the mass of some heavy scalar fields in the adjoint representation). Then we would have a phase diagram as in the Figure below.

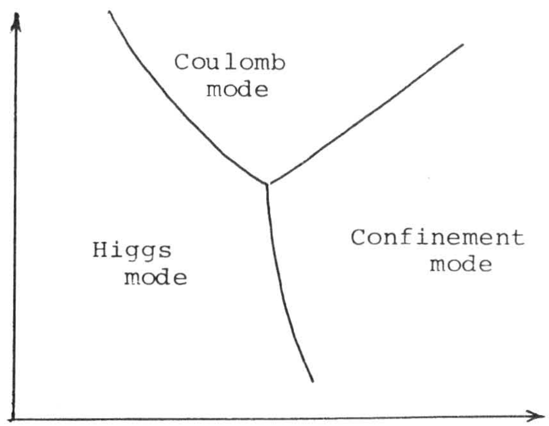


Fig. 1

Numerical calculations[20] suggest that the phase transition between the two confinement modes is a first order one. Real QCD is represented by one point in this diagram. Where will that point be? If it were in the Coulomb phase there would be long range, strongly interacting Abelian gluons contrary to experiment. In the Higgs mode quarks would have finite mass and escape easily. It could be still in the Higgs phase but very close to the border line with the confinement mode. If the phase transition were a second order one then that would imply long range correlation effects requiring light physical gluons. Again, they are not observed experimentally. If, wich is more likely, the phase transition is a first order one then even close to the border line not even approximate confinement would take place: quarks would be produced copiously. There is only one possibility: we are in the confinement mode. Electric flux lines cannot spread out. Quark confinement is absolute.

REFERENCES

1. S. Weinberg, Phys. Rev. Lett. 19 (1967) 1264.
A. Salam and J. Ward, Phys. Lett. 13 (1964) 168.
S.L. Glashow, J. Iliopulos and L. Maiani, Phys. Rev. D2 (1970) 1285.
2. E.S. Abers and B.W. Lee, Physics Reports 9C no. 1.
S. Coleman in: Laws of Hadronic matter, Erice July 1973, ed. by A. Zichichi, Acad. Press, NY and London.
3. G. 't Hooft, lectures given at the Cargèse Summer Institute on "Recent Developments in Gauge Theories", 1979 (Plenum, New York, London) ed. G. 't Hooft et al. Lecture no II.
4. H. Georgi and S.L. Glashow, Phys. Rev. Lett. 28 (1972) 1494.
5. G. 't Hooft, in "The Whys of Subnuclear Physics", Erice 1977, ed. by A. Zichichi (Plenum, New York and London) p. 943.
6. G. Parisi, lectures given at the 1977 Cargèse Summer Institute.
7. P. Higgs, Phys. Rev. 145 (1966) 1156.
T.W.B. Kibble, Phys. Rev. 155 (1967) 1554.
8. S. Coleman, in "New Phenomena in Subnuclear Physics", Erice 1975, Part A, ed. by A. Zichichi (Plenum press, New York and London).
9. G. 't Hooft, in "Particles and Fields", Banff 1977, ed. by D.H. Boal and A.N. Kamal, (Plenum Press, New York and London) p. 165.
10. P.A.M. Dirac, Proc. Roy. Soc. A133 (1931) 60; Phys. Rev. 74 (1948) 817.
11. R.F. Streater and A.S. Wightman, PCT, spin and statistics, and all that (Benjamin, New York and Amsterdam, 1964).
12. H. Lehmann and K. Pohlmeyer, Comm. Math. Phys. 20

(1971) 101;
A. Salam and J. Strathdee, Phys. Rev. D1 (1970) 3296.
13. G. 't Hooft, Nucl. Phys. B138 (1978) 1.
14. L.P. Kadanoff and H. Ceva, Phys. Rev. B3 (1971) 3918.
15. M.B. Nielsen and P. Olesen, Nucl. Phys. B61 (1973) 45; ibidem B160 (1979) 380.
16. B. Zumino, in "Renormalization and Invariance in Quantum Field Theory", ed. E.R. Caianiello, (Plenum Press, New York) p. 367.
17. G. 't Hooft, Nucl. Phys. B153 (1979) 141.
18. R. Jackiw and C. Rebbi, Phys. Rev. Lett. 37 (1976) 172.
C. Callan, R. Dashen and D. Gross, Phys. Lett. 63B (1976) 334 and Phys. Rev. D17 (1978) 2717.
19. G. 't Hooft, Nucl. Physics B79 (1974) 276
A.M. Polyakov, JETP Lett. 20 (1974) 194.
20. M. Creutz, L. Jacobs and C. Rebbi, Phys. Rev. D20 (1979) 1915.